

**National Assessment of Educational Progress (NAEP)
Technology-Based Assessment (TBA) Project
Writing Online (WOL) Pilot Test Field Notes
March 21, 2001**

Background

The Writing Online (WOL) project is one of three empirical studies currently under way at ETS in association with the NAEP Technology-Based Assessment (TBA) project. The TBA project is intended to lay the groundwork for incorporating new technology into NAEP.

Research Questions

The TBA project is envisioned to address the following research questions:

- *What are the measurement implications of using TBA in NAEP?*
- *What are the equity implications?*
- *What are the efficiency implications of using TBA compared with paper and pencil assessment?*
- *What are TBA's operational implications?*
- *How can we best incorporate new technology into NAEP in both the short and long term?*

Specifically, the WOL study is designed to investigate the following questions:

Measurement

- How does test mode (i.e., computer v. paper) affect the inferences drawn about students' writing skill?
- Do students perform differently across the two modes?

Equity

- How are students with different levels of computer experience impacted by technology v. paper-based writing assessment?
- How do population groups perform and, particularly, do mode effects vary across groups?
- How might we use computer delivery to accommodate the needs of students with disabilities?

Efficiency

- Is a technology-based writing assessment more cost effective or timely than a paper-based one?
- Can web delivery and automated essay scoring reduce costs and shorten reporting lag times?

Operational feasibility

- What are the logistical challenges associated with administering a NAEP survey on computer?
- Are school facilities, equipment, software, and Internet connectivity, administrator effectiveness, school cooperation, and data quality sufficient to the task?

WOL Pilot Test

The WOL study will consist of three phases: pilot test, pretest, and field test. Pilot testing for WOL was conducted between November 6 and November 14, 2000. The pretest was scheduled for April to May 2001 and the field test for March to May 2002. Each of these phases was intended to serve a different purpose. The pilot test, which is the subject of this report, was conducted to assess software functioning and logistics. The pretest was conducted to assess how effectively our software and logistics scaled up. Finally, the field test will address the main study questions listed above.

Description of the Sample

A convenience sample was selected for the pilot testing, which was conducted at five New Jersey schools located in Ewing Township, Long Branch, Middletown Township, Pennington, and Ringoes.

A total of 51 students (46 at eighth grade and 5 at seventh grade) participated in the pilot testing. The breakdown of students by gender was 25 males and 26 females.

In addition, WOL was administered to 12 students in Australia via the Internet. They experienced no difficulties in testing and were able to electronically return their essays to ETS for scoring. The principle we demonstrated was that, via the web, we can administer assessments anywhere in the world, with minimal or no intervention by our staff.

Hardware and Software Requirements

The following hardware and software were required to administer the WOL test:

Hardware

- Internet connection: dedicated (non-dial up) 128Kb per second or greater line
- Computers: PC with Pentium class 166 megahertz microprocessor or better
- Memory: 32MB or greater
- Operating system: Windows 95, Windows 98, Windows Me or Windows NT
- Hard drives: 10MB free disk space required
- Graphics capabilities: SVGA support—800x600 resolution (or better) with minimum 256 colors

Software

- Web browser: Microsoft's Internet Explorer Version 5

Some minor enhancements to Internet Explorer were also required. These were installed during the certification process if they were not already on the computer. The enhancements included:

- Internet Explorer core fonts and additional web fonts
- Macromedia Flash 4.0 plug-in
- Microsoft Virtual Machine (Java)

Description of Instrument

The test instrument consisted of four parts: tutorial, computer skills measure, writing test, and background questions.

Tutorial

The tutorial asked students to indicate whether they knew how to use the mouse and to scroll. Then, depending on their responses, the tutorial either gave them further instruction or a short test exercise. No students took the “No” path, that is, all indicated that they knew both how to use the mouse and how to scroll. The tutorial also showed the layout of the test screens, indicating where students would find word processing tools and where they should type their essays. All of the word processing tools were introduced and defined, then students were given two minutes to practice typing and to try out the word processing tools (cut, copy, paste, and undo).

Computer Skills Measure

The computer skills measure consisted of a test of keyboarding speed and accuracy and five editing tasks. For the keyboarding exercise, students were shown a paragraph and told to type it exactly as written. They had two minutes to type and were moved on automatically at the end of the two-minute period, if they had not finished within that timeframe. Students were not allowed to use the word processing tools during this exercise.

For the editing tasks, students were shown a paragraph and were instructed to make five specific changes to the text. The editing tasks, which were shown on five separate screens, included correcting a misspelled word, inserting a word, deleting a word, changing a word, and moving text. Students were given one minute for each of the tasks, and they were moved on automatically to the next task at the end of the one-minute period, if they had not finished within that timeframe. Students were allowed to use the word processing tools during this exercise.

Writing Test

The writing test consisted of two 25-minute essay prompts—one informative topic and one persuasive topic. Students were shown general directions on a screen, then they moved to the first essay prompt (informative). The prompt was displayed on the left side of the screen, and students typed their essay in a field on the right side of the screen. After 25 minutes elapsed, they were moved automatically to the second essay prompt (persuasive). The layout for the second prompt was the same as described above.

Even if students finished typing their essays before 25 minutes had elapsed, they had to wait for the full 25 minutes on each essay. Many were unhappy with that situation, and some students began to experiment with the keyboard, testing out what would happen when they highlighted, clicked, and moved around the screen. When asked what they would like to change about the test, many students indicated that they wanted to move on when they were finished, whether or not the 25 minutes had elapsed.

Background Questions

The background questions section consisted of 34 questions: 9 NAEP general background questions (including race/ethnicity, parent education level, and home literacy factors) and 25 questions about students' experience with computers. Students were shown the question in a window on the left of the screen, and the response options were displayed in a window on the right of the screen. They were directed to click on the bubble next to their selected response. Students were able to move ahead or back throughout this section by clicking on the "Next" and "Previous" buttons. A counter in the upper-right corner of the screen indicated which question they were answering, for example, "27 of 34 questions."

Students had 20 minutes to complete this section, but most of them completed all of the questions in about 7 to 10 minutes. Only two or three of the students needed any clarification or asked any questions during this section. Several students, in fact, indicated that answering these questions was "fun" because it was about them and their experiences.

Remote PC Certification

School staff members were asked to verify that their PCs had the hardware and software capabilities needed to administer WOL. This verification was done via a remote certification process. The process required schools to log on to a web site and follow instructions on the computer screen, which allowed our software to scan their PCs. Two schools were able to certify at least a few of their computers via the remote process, but most needed assistance from ETS staff, either on site or by telephone, to complete their certification.

Although some computers "failed" certification remotely, they were able to be certified on site and could deliver the test satisfactorily. Also, several computers did not meet the requirements for delivering the tutorial, but were able to satisfactorily deliver the actual test. Because of this situation, the minimum throughput threshold required by the certification process needs further discussion, as we do not want to unnecessarily fail any computers due to low throughput.

Test Delivery and Administration

Students took the test via web delivery, using their own desktop computers, or via laptops brought into schools by NAEP staff. Laptops were not connected to the Internet, but ran the testing software locally. One school tested entirely via web delivery, one school

tested entirely via laptop delivery, and three schools used a combination of laptop and web delivery. Overall, 31 students tested via the web and 20 students tested on laptops.

Initially, we encountered some difficulty with firewall or security software in three schools. However, we were able to make the necessary adjustments, with the assistance of ETS technical staff on site or by working with school technical staff over the phone. Generally, school staff members had to make modifications to allow the downloading of browser extensions and to decrease the security levels so our web site was accessible. We experienced no server breakdowns or crashes, which may have been due, at least partially, to having no more than six students testing simultaneously via the web. Throughput remained satisfactory at all the schools throughout the administrations. Students experienced little latency, or lag time, when moving from one screen to another.

School staff members and officials were cooperative and worked diligently to meet all of the WOL equipment requirements. As a result, there were no major technical problems at any of the schools. During the first two administrations, however, there were some minor technical difficulties:

- The tool bar expanded and contracted when students clicked on it;
- Ovals on the background questions did not immediately appear when the question screens came up; and
- The animation of the runner in the tutorial did not work if students did not click exactly on the numbers in the center of the screen.

All of these problems were resolved on site by ETS staff.

At each school, the administrator logged students onto the computer through an administrator screen that required a password, school ID number, and student ID number. Once logged on, students received all of their instructions and completed all of their work on the computer. When the final screen came up, indicating that the test was finished, students raised their hands to inform the administrator they were done. After a few verbal feedback questions (see Student Feedback), students were dismissed. The administrator then logged off the computer, or returned to the login screen to prepare for the next student.

Within schools where both laptops and desktops were used, students were randomly assigned to the machines. Although only two or three students indicated that they had ever used a laptop before, students were generally excited about the opportunity to use the laptop. Rather than the built-in track ball pointer, students used an external mouse with the laptop, which was probably more familiar to them. Although students appeared to exhibit few problems adapting to the laptop, we do not know if they would have written longer or better essays had they been able to use full-size keyboards. Two students did comment that the keyboard was a bit smaller than they were used to, but they seemed to type their essays without too much difficulty, and one student said he would have preferred a “regular” keyboard because he was “used to that.”

By trying a combination of keystrokes, several students who were tested on the laptops were able to move out of the essay section before 25 minutes had elapsed. This will not be a problem in the pretest, however, as we will allow students to move ahead when they finish their essays.

Each student received a brochure titled “Ideas for planning and reviewing your writing,” which was used in the 1998 NAEP writing assessment and will also be used in the 2001 writing field test. Students could refer to the brochure at any point during the assessment, but they were specifically instructed to look at it prior to beginning their writing tasks. Only a small number of students looked at the brochure.

Two staff members from Westat—the NAEP sampling and data collection contractor, observed one of the administrations to become more familiar with the test and to offer suggestions regarding operational logistics.

Word Processor

Students generally reacted favorably to the NAEP word processor. They found it easy to use and very much like the word processors they were accustomed to using at home or in school. Interestingly, many students did not make use of the word processing tools. For example, most students deleted text simply by backspacing, rather than highlighting and clicking on the Cut icon, or they deleted text and retyped it, rather than using the Cut and Paste tools. Also, although interface tools for underlining and boldface were not part of the NAEP word processor, a number of students made use of shortcut keys for those functions (e.g., Ctrl-U for underlining).

Because the spellchecker was not bug-free by the time of the administrations, it was disabled for the pilot test. We did, however, inquire of students whether they use the spellchecker when they write on the computer. All of them indicated that they use the spellchecker, generally running it on the whole piece after they are finished writing or after they write a few paragraphs. A few students also commented that they use the grammar checker, which was not available on our word processor.

Scoring

Scoring the Word Processing Measure

A program that implements a version of the Boyer-Moore (B-M) string-searching algorithm, considered one of the most efficient string-matching techniques, scored the word processing measure. The algorithm scans the characters of the pattern (in this case, text the student retyped exactly from a paragraph we presented) from right to left beginning with the rightmost position. The errors were categorized as: capitalization, punctuation, spacing, and major or minor typographical errors (dependent on the number of missing or juxtaposed characters). Results of the word processing measure will be used to begin development of scales for sorting students into word processing skills or ability categories.

Manual Essay Scoring

The pilot study scoring consisted of 56 responses to the informative prompt and 53 responses to the persuasive prompt. Human raters scored the essays on a 6-point scale, using the scoring guide for the NAEP writing assessment. The raters scored all responses twice.

Student Feedback

After completing the test, students were verbally asked several questions about their reactions to testing on the computer:

- What did you like or dislike about the test?
- Do you prefer writing your essays on the computer or with paper and pencil?
- Do you prefer the essay prompt to the left of the typing area or horizontally above the typing area?
- What suggestions do you have for improving the test?

Overall students were very positive about the test. Many commented that the essay prompts were interesting and easy to write about, directions were easy to follow, showing time elapsed was helpful, and background questions were interesting. Almost all of the students indicated that they prefer to type their essays on computer, rather than writing them by hand. The most common reasons cited were that it is easier to make corrections on the computer, and they find it less tiring than handwriting. Some students also said that they have poor handwriting, and typing their work makes it legible. A few students also indicated that they like to do their prewriting on paper, then type their essay.

The majority of the students said that they preferred the essay to the left of the typing area, as it was natural to read on the left and work on the right side of the screen. By far, the two changes students would like to make to the test are being able to move ahead before the 25 minutes elapses on the essay and having a spellchecker available. Almost every student who commented mentioned either one or both these changes.

Next Steps

- Make spellchecker tool active for all students. Students almost unanimously said that they use spellchecker when they write on the computer. Their use of such a tool suggests that an authentic computer-based writing assessment should include a spellchecker tool. At the time of the pilot, the spellchecker tool was not activated because it was not yet fully developed. The spellchecker tool will be available for all students during the pretest in spring 2001.
- Add description of Hide/Show button to the tutorial. The Hide/Show button, which hides the essay prompt and displays a larger typing area, was activated for the pilot but was not covered in the tutorial. Although only a small number of students tried it, many others indicated that they would have used it if they had understood what the button meant.

- Add description of Previous and Next buttons and the time clock to tutorial. Few students had any questions about or difficulty using the Previous and Next buttons; however, the tutorial will be more complete if all buttons are explained. Also, some students did not notice the time clock in the upper-left corner of the screen, which is particularly important when working on the essays. We will revise the tutorial to include the time clock, largely to bring it to students' attention.
- Investigate further whether to use external keyboards with laptop computers. Because it will likely be necessary for a number of students in the pretest to use laptop computers, and because few students are likely to have previous experience using the laptop keyboard, we need to investigate further any impact of using the laptop versus the regular computer keyboards. Within schools, we will compare the essays and scores of students who tested on the laptop with those of students who tested using the regular computer keyboard in order to identify whether there are any noticeable patterns that may be attributed to using the laptop, for example, of keyboarding errors or essay length. If necessary, we will discuss with Westat the need to use external keyboards with the laptops.
- Continue to refine remote PC certification process. Based on feedback from Westat (NAEP sampling and data collection contractor), we have already made some revisions to the remote PC certification screens and instructions. We will continue to refine the process as we move toward the spring 2001 pretest. Also, our minimum throughput threshold needs further discussion because we found that some computers that failed remotely could be certified on site and were able to deliver the WOL test satisfactorily.
- Develop online essay-scoring interface. To facilitate the scoring process, we will develop an online essay-scoring interface that will allow scorers both to view the student responses and to enter their scores online.